What do malware analysts want from academia? A survey on the state-of-the-practice to guide research developments

Marcus Botacin[1]

[1]Assistant Professor
Texas A&M University (TAMU), USA
botacin@tamu.edu
@MarcusBotacin

Agenda

Agenda

Literature Review

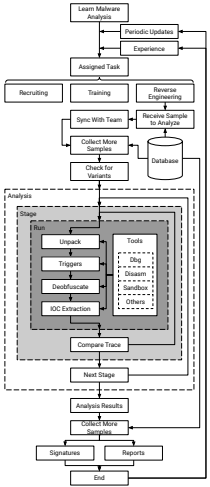
Table: **Paper Selection.** Paper distribution per year (2000 – 2018) and per venue for the Original and the Refined SLR.

Venue/Year	0		1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		16		17		18		Total	
	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R	O	R
USENIX	1	0	0	0	0	0	0	0	0	0	1	0	1	0	6	2	2	0	3	1	7	1	8	1	10	1	12	0	9	2	7	0	9	3	13	1	6	0	95	12
CCS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	4	1	6	0	6	0	7	0	11	0	9	2	11	1	14	0	2	0	11	2	6	0	89	6
ACSAC	0	0	0	0	0	0	0	0	2	0	3	2	2	0	4	0	4	1	1	0	3	0	8	0	10	3	7	0	10	0	6	1	3	1	7	0	8	0	78	6
IEEE S&P	0	0	1	0	0	0	0	0	0	0	1	0	3	2	2	1	1	0	0	0	0	0	10	0	17	2	12	0	3	0	6	1	4	2	5	1	3	1	68	11
DIMVA	0	0	0	0	0	0	0	0	0	0	4	1	4	0	3	0	8	0	2	0	3	0	0	0	8	1	4	1	8	1	7	0	7	2	5	1	4	2	67	9
NDSS	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	3	0	3	1	3	1	3	0	2	0	4	0	5	0	4	1	9	1	7	0	3	1	49	5
RAID	0	0	0	0	1	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	3	0	5	1	5	1	3	0	4	1	3	0	3	0	31	3
ESORICS	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	2	1	1	0	0	0	0	0	2	0	3	0	3	0	0	0	1	0	1	1	0	0	14	3
Total	1	0	1	0	1	0	0	0	3	0	11	4	15	2	17	3	24	3	16	2	22	2	36	1	63	7	56	4	54	5	47	3	39	10	52	6	33	4	491	55

Table: 2

Table: **Original SLR Source:** "Challenges & Pitfalls in Malware Research - Botacin at al."

Systematization



Knowledge Gaps

1. How do they learn?
2. Which threats do they analyze?
3. What are the analysis practices?
4. How do they report results?

Agenda

Methodology

- Online survey with 30 questions (IRB-approved).
- **Initial run with 21 analysts (confirmed professionals).**
- Replication study with another 21 analysts (from the Web).

The Analysts and the Malware Analysis Job

The Participants

Table: **Analysts' occupation.**

#	Role	Company	Obs.
1	CISO	Non-Security	
1	Threat Hunter	Intelligence Agency	
1	Leader	Government CSIRT	
1	Member	Bank CSIRT	
4	Consultant	Independent	Ex AV analysts
5	Analyst	Sec. Consultancy	2 companies
8	Analyst	AV company	4 AV companies

The Job

Table: **Analysts' Malware Analysis Tasks Frequency.**

Category	Full	Most	Reasonable	Eventual	Never
Answers	3 (14%)	5 (24%)	6 (29%)	7 (33%)	0 (0%)

Table: **Analysts' Type of Tasks vs. Analysis Teams.**

Category	Team/Together	Team/Individual	Independent
Answers	1 (5%)	16 (76%)	4 (19%)

The Knowledge

Table: **Analysts' Strategies for Learning Malware Analysis.**

Category	Post-Grad	Major	Cert.	Work	Self
Answers	2 (10%)	0 (0%)	0 (0%)	10 (48%)	9 (42%)

Table: **Analysts' Knowledge Updating Strategies.**

Category	Academic Papers	White Papers	Videos	Events	Training
Answers	15 (71%)	21 (100%)	13 (61%)	18 (85%)	12 (57%)
Rate	14%	46%	11%	21%	12%

The Expertise

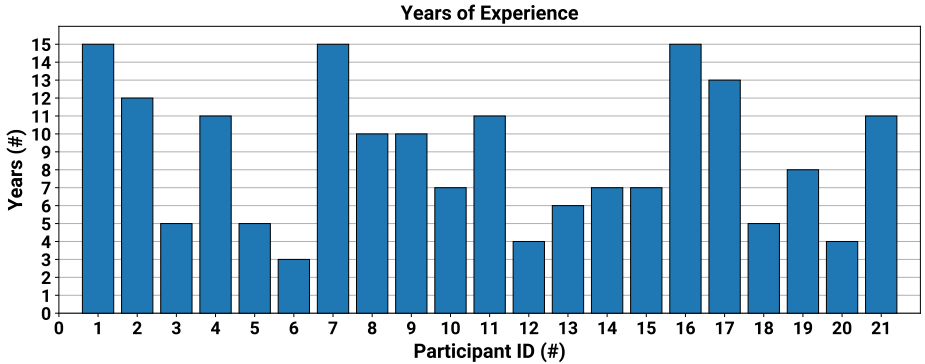


Figure: **Years of Experience.**

The Analysis Practices

The Samples

Table: **Knowledge on Samples' Collection Context.**

Category	Regional	Local	Unknown
Answers	11 (52%)	3 (14%)	7 (33%)

Table: **Additional Samples Collection by Analysts.**

Category	Sig.	Report	Understand	No
Answers	8 (38%)	2 (10%)	7 (33%)	4 (19%)

The Variants

Table: **Samples storage**

Category	OC	OA	SC	N
Answers	9 (42%)	5 (24%)	5 (24%)	2 (10%)

Table: **Malware Variants Re-Analysis Rate.**

Categories	Very Often	Sometimes	Rare
Answers	11 (52%)	9 (43%)	1 (5%)

The Tools

Table: **Analysis hosting.**

Category	Own	Public	Company
Answers	18 (85%)	1 (5%)	2 (10%)

Table: **Use of public sandboxes.**

Category	Like	Dislike	Disallow
Answers	11 (52%)	6 (28%)	4 (20%)

The Automation

Table: **Analysis Automation Rates.**

Category	Fully	Half	Manual
Answers	0 (0%)	11 (52%)	10 (48%)

Table: **MultiStage Handling.**

Categories	MA	MI	MM	FA	SS
Answers	1 (5%)	14 (66%)	4 (19%)	2 (10%)	0 (0%)

Time & Skills

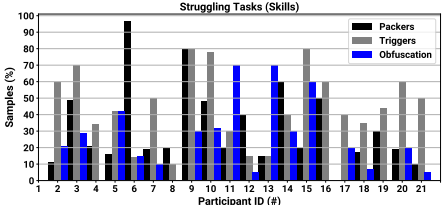


Figure: **Analysts' Most-Struggling Tasks (Skill-Wise).**

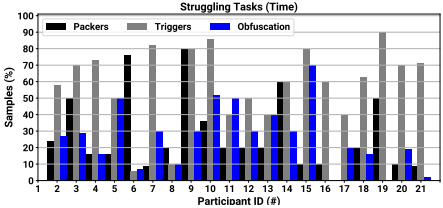


Figure: **Analysts' Most-Struggling Tasks (Time-Wise)**

The Analysis Accuracy

Tools Selection

Table: **Number of Typical Analysis Runs.**

Category	A1	SM	TC	AC
Answers	0 (0%)	8 (38%)	8 (38%)	5 (24%)

Table: **The Use of Different Sandboxes by Analysts.**

Category	A1	SM	TC	AC
Answers	1 (5%)	8 (38%)	9 (42%)	3 (15%)

Multi-Path Samples

Table: **Environment Configuration by the Analysts.**

Category	Both	Arch	OS	None
Answers	5 (24%)	2 (10%)	0 (0%)	14 (66%)

Table: **Most-Used Path Exploration Strategies.**

Category	Fuzzing	Symbolic	Concolic	Forced	Manual
Answers	9 (42%)	7 (33%)	5 (23%)	14 (66%)	19 (90%)
Rate	35%	41%	29%	49%	73%

Comparison & Validation

Table: **Most-Used Trace Comparison Strategies.**

Category	All Traces	IoCs	Graphs
Answers	6 (28%)	13 (62%)	2 (10%)

The Analysis Tools

Most Used Tools

Table: **Tools Usage.**

Category	Similarity Hash	Debugger	Sandbox	Decompiler	Unpacker	AntiVirus
Answers	16 (76%)	18 (86%)	20 (95%)	19 (90%)	19 (90%)	11 (52%)
Rate	47%	57%	58%	61%	49%	58%

Table: **Analysts' Perception about Debuggers Usefulness.**

Category	Repetitive	Enough	Not essential
Answers	15 (71%)	4 (19%)	2 (10%)

Most Helpful Tools

Table: **The Role of Debugger Plugins for Malware Analysis.**

Category	Essential	Specific	No Difference
Answers	9 (42%)	12 (48%)	0 (0%)

Table: **The Role of Decompilers in Malware Analysis.**

Category	Very	Minor	Not Useful
Answers	17 (81%)	4 (19%)	0 (0%)

Performance Considerations

Performance of Analysis Tools

Table: **Analysts' Perception About Tools Performance.**

Category	SF	SI	FE
Answers	10 (47%)	3 (15%)	8 (38%)

Table: **The Usefulness of Faster Sandboxes.**

Category	Very	Specific	No Diff
Answers	10 (48%)	11 (52%)	0 (0%)

Performance of Matching Tools

Table: **Most-Frequent Analysis Outcomes.**

Category	Both	Reports	Signatures
Answers	10 (48%)	9 (42%)	2 (10%)

Table: **Required Properties for Signature Generation.**

Category	Same	Acc. First	Only Acc.
Answers	7 (33%)	10 (47%)	4 (20%)

The Future Tools

Engineering Developments (1/2)

Analysts want more scalability:

P13. "A Windows VM provided by Microsoft without many security things and tailored to allow me to change any characteristics of the machine without much trouble, like language, username, etc."

Analysts want better Usability:

P18. "Better GUI based API tracer (similar like outdated API monitor)"

P8. "I wish x64dbg could be called from the CLI and run a script with a sample."

Engineering Developments (2/2)

Analysts want more efficiency:

P8. "In Linux, I'd like to have more injection capabilities in strace and a Yara-like tool to match instructions."

Analysts want to increase accuracy:

P6. "Better Unpackers."

Scientific Developments (1/2)

Analysts want increased accuracy:

P5. "Multi-Architecture Sandbox."

Analysts want more usability:

P7. "A more automated angr."

P8. "A good API logger that doesn't require me to choose which function calls I want to see. Something like strace but for Windows."

P14. "A memory monitoring tool that you attach to a process before executing it and it automatically dumps anything interesting..."

P20. "AI-assisted function identification for stripped binaries that actually works."

Scientific Developments (2/2)

The case of decompilers

P9. "better decompilers to languages like delphi, go, rust."

P14. "An easy-to-use decompiler based on the execution trace (for virtualized samples)"

P19. "Improved decompilers with better types and static library detection; better ways to identify malware families."

P10. "IA behavior analysis based on intermediate machine code."

The future of malware analysis

Summary

Table: **Analysts' Impressions about AI usage.**

Category	Solve	Help	No Change
Answers	1 (5%)	19 (90%)	1 (5%)

Moving Forward

1. **From analysts to intelligence.**
2. **AIs will not replace humans.**
3. **We need better education of human analysts.**

Analysts Opinions (1/3)

Malware Tactics:

P5. "It will require more and more skilled people. Malware evasion are common place now."

P7. "Multi-stage, fileless, firmware and other types of samples that are difficult to analyze with traditional techniques will have a great impact on users' security, but at the same time will provide new opportunities for research in the field."

P5. "With the increase of ARM devices, I believe we will have an increase of multi-architecture malware (recently I have seen an increase of multi-platform malware, but multi-architecture is still rare)."

Analysts Opinions (2/3)

Developing Intelligence:

P15. "Focus will change from file/code analysis on initial attack vectors (phishing, social engineering, network behavior etc.)"

P8. "Being able to fully analyze a malware sample/family is not the most important thing IMHO. We have to have context and we need to extract intelligence from it, not only describe its features. Maybe we have to interact with its C2, track the actors, etc. So, malware analysis plays a key part on campaign/incident investigation, but it doesn't help much alone."

Analysts Opinions (3/3)

The role of AI:

P12. "I think the presence of a malware analyst will always be necessary. Perhaps there will be a day when an AI will be able to analyze with precision, but even in this case there will have to be a malware analyst to "feed" the AI with more inputs and progress the techniques and tools."

P13. "AI will help and eliminate trivial tasks, but often is necessary to perform advanced tweaks to make the malware work, So, this needs to be done by a human being."

P14. "AI will be useful for anomaly detection, but manual malware analysis will still be required to better understand how the attack works."

P18. "AI will help in future more but there'll be always a need for analysts."

Moving Forward

Education:

P19. "We need better education, but it is a niche job."

P1. "An ever-growing field with a great need for great and open-minded researchers. Start to think like attackers and combine it with the mindset of a defender and you'll more chance to win."

P9. "Always will raise new challenger malware that will need skilled professionals and better courses will be a differential to prepare new professionals."

Agenda

Moving Forward Summary

#	Finding	Suggested Direction
1	Malware analysts perform more and varied daily tasks than reverse engineering all day.	Develop tools that allow easy context switching.
2	Most malware analysts work in teams, but they analyze samples individually.	Develop collaboration tools that focus more on the sharing of the final result than on real-time collaboration.
3	Most analysts have to handle regional threats.	Develop more region and context-specific malware evaluations, such as region-specific longitudinal studies.

Moving Forward Summary

#	Finding	Suggested Direction
4	Most professionals are self-taught malware analysts.	Develop more malware courses in the universities.
5	Reading papers is the preferred form of getting updates for most analysts. However, most analysts read more white papers than academic papers.	Make academic papers reach out to professional communities to increase their impact and better support security professionals.
6	Most analysts collect additional samples to enrich their analysis procedures.	Enhance similarity detection tools for threat triaging.

Moving Forward Summary

#	Finding	Suggested Direction
7	Most malware analysts still receive recognizable malware variants for analysis.	Enhance similarity detection tools for threat triaging.
8	Many analysts end up hosting their own analysis solutions rather than using a COTS one due to their lack of configuration possibilities.	Service-based solutions such as public sandboxes should be more customizable.
9	Some analysts use their own analysis solutions due to companies not allowing the use of public services.	Develop easier-to-install and easier-to-configure solutions to not put the configuration burden on the analyst.

Moving Forward Summary

#	Finding	Suggested Direction
10	Most analysts still handle multi-stage malware via multiple, non-integrated tools.	Increase the integration between tools, such as via standardized data transfer protocols
11	Most analysts still handle multi-stage manually.	Develop automation tools that integrate different types of threats, and not only support different tasks for the same threat type.
12	Unpacking samples is hard, regardless of the malware analyst's expertise level.	Develop automated unpacking and obfuscation tools.

Moving Forward Summary

#	Finding	Suggested Direction
13	Unpacking and deobfuscation are also time-consuming, even for skilled analysts.	Develop automated unpacking and obfuscation tools.
14	Most analysts do not run analyses multiple times or in multiple sandboxes as a standard practice.	Develop guidelines and metrics to evaluate when a sample requires additional inspection.
15	Most analysts explore multiple execution paths manually and not via structured approaches and solutions described in the literature.	Popularize solutions for automatic multipath exploration such as fuzzing and symbolic execution.

Moving Forward Summary

#	Finding	Suggested Direction
16	Half of all surveyed analysts believe that the performance of analysis solutions can be improved.	Develop faster sandboxes, that are acknowledged by most analysts as a point of improvement.
17	Decompilers are the most useful tool in most analysts' opinion even though decompiler limits are widely acknowledged by them.	Develop more decompilers focused on malware analysis because, despite decompiler limits, it is the tool that most helps analysts.
18	An increased automation level for the analysis tools is desired by most analysts.	Benefit from AI developments to create automated hooking and automation function identification tools.

Moving Forward Summary

#	Finding	Suggested Direction
19	Most analysts believe AI will help in their work, but they believe analysts are still required to train the AI models.	Train new analysts in the creation of AI-assisted security solutions and the creation of security core knowledge for these solutions.
20	Education is voluntarily pointed out by most analysts as the most required change for the future.	Focus on the training of the next generation of malware analysts workforce with special attention in the development skills to understand attacker's mentality.

Thanks!

Questions? Comments?

botacin@tamu.edu
@MarcusBotacin